STAT 830 Lecture 3
Fall 2020
17 September 2020

- We have finished "Nonparametric Basics".
- On Tuesday next week I will talk about approximate bias and variance of a ratio estimate and then discuss the bias-variance trade-off in a nonparametric we will continue "Nonparametric Basics".
- Then I will start on the "Likelihood Basics" module if I have time.
- Next Thursday I promised a quiz. I think that must be postponed to Thursday Oct 1.
- In the meeting today I forgot, again, to record the session; I am begging students to remind me next time.
- Today I talked about bootstrapping and Monte Carlo. I used the `R` code here.
- The basic bootstrap is to approximate the population distribution of a random quantity by the Monte Carlo distribution of the same random quantity when the population is replaced by the sample distribution. We simply draw an iid sample of size n from the original n data points.
- I investigate how well the bootstrap works in one small problem: getting Confidence Intervals for the population mean μ.
- I used two methods: $\bar{X} - Q_2, \bar{X} - Q_1$ is one interval (labelled z in the `R` coode; and $\bar{s} \pm cs/\sqrt{n}$. The bootstrap is used to find good values of Q_1 and Q_2 and of c.
- If $\bar{X} - Q_2, \bar{X} - Q_1$ is a confidence interval with coverage probability $1 - \alpha$ then
$$P(\bar{X} - Q_2 < \mu < \bar{X} - Q_1) \approx 1 - \alpha$$
which is the same as
$$P(Q_1 < \bar{X} - \mu < Q_2) \approx 1 - \alpha.$$
If we knew the distribution of $\bar{X} - \mu$ we could calculate a suitable Q_1 and Q_2 so that
$$P(\bar{X} - \mu < Q_1) = \alpha/2 = P(\bar{X} - \mu > Q_2).$$
We don't know that distribution but we hope it is close to the distribution of
$$\bar{X}^* - \mu^*$$
where the X^* values are sampled with replacement from the original X values and $\mu^* = \bar{X}$ is the mean of the sample (which is the population mean for the bootstrap samples).

We do know, from our bootstrap samples the distribution of $\bar{X}^* - \mu^*$; so we can find Q_1 and Q_2 which work for this bootstrap distribution.

- We also pretend that the distribution of

$$\frac{\bar{X} - \mu}{s/\sqrt{n}}$$

is close to that of

$$\frac{\bar{X}^* - \mu^*}{s^*/\sqrt{n}}$$

and this distribution can be found, approximately, from our bootstrap samples.

Then you use this to find c and produce t intervals with the multiplier computed this way and *not* from t-tables.

- I studied these approximations in a Monte Carlo sample. For 2 distributions (Uniform[0,1] and standard exponential) I generated 100 samples of size $n = 5, 25$. For each sample I used bootstrapping to compute the two confidence intervals above. Then I counted up the fraction which covered the mean of the uniform (0.5) or exponential (1) distribution in question.

- For sample size 5 the coverage probabilities are not very close to 0.95, particularly for the exponential case. They are better for sample size 25. But this Monte Carlo study had only 100 replicates due to computing times so the coverage probabilities are not estimated very accurately.

- I also pointed out the effect of asymmetry in the exponential case where the intervals were much more likely to miss on the low side.

- Handwritten slides.